

Curso de verano de la Universidad de Navarra.

La Hoja de ruta de los archivos personales, familiares y de empresa

Recomendaciones para la digitalización de documentación manuscrita

Creación, conservación y difusión de Archivos digitales.

L.M.Daniel Iturri Sesma

Técnico en digitalización del Centro de Patrimonio Documental de Euskadi,
IRARGI, del Gobierno Vasco.

0. Introducción.

1. Conceptos digitales básicos.

1.1. Imagen digital.

1.2. Resolución.

1.3. Profundidad de color.

1.4. Formatos de archivo.

1.5. Soportes de archivos digitales.

2. Tamaños y posicionamiento de la documentación en la toma de imagen.

3. Creación del Registro/Archivo digital.

3.1. Recomendaciones para la creación del Registro/Archivo digital original como
archivo de seguridad y conservación.

3.2. Recomendaciones para las copias de difusión.

4. Fases del proceso de digitalización y control de calidad.

0. Introducción

La digitalización es una técnica reciente todavía sometida a diferentes corrientes de mercado en su afán de posicionarse como solución definitiva. No existe un consenso sobre técnicas o normas (ISO, UNI, ANSA, DIN, AFNOR, UNE) estandarizadas para su ejecución, por ello, reina una sensación de anarquía entre las Organizaciones y empresas de servicio que trabajan en este ámbito.

Hasta hace poco la oferta de equipos que ofrecieran prestaciones adecuadas a las demandas era muy escaso y a veces no ofrecían los resultados deseados.

La evolución más significativa aparecida en estos últimos años ha sido la incorporación de las cámaras digitales fotográficas en sustitución de los escáneres. Esto ha permitido un fuerte acelerón en la productividad del trabajo y se están imponiendo como periféricos idóneos para esta fin.

El objetivo principal de la digitalización de un documento manuscrito es ofrecer al usuario una imagen del documento original, con una calidad visual que permita prescindir del acceso al documento original. Esto revierte en una mejor conservación del mismo, evitando el posible deterioro debido a múltiples manipulaciones y, como objetivo secundario, no menos importante, permitir un rápido y cómodo acceso al contenido de documento.

Desde esta perspectiva se va a tratar de dar unas Recomendaciones para:

- normalizar el proceso y la terminología
- proponer unos parámetros técnicos
- describir un protocolo de actuación para la creación de Registros/Archivos digitales originales y, a partir de ellos,
- crear archivos para difusión.

1. Conceptos digitales básicos

1.1 Imagen digital.

Cuando se contempla de cerca un mosaico romano, se observa que está formado por pequeñas piezas de cerámica coloreadas, llamadas teselas, que delinean y dan color a las formas de la imagen que representa.

El sensor que incorpora la cámara fotográfica digital, al igual que el mosaico, esta compuesto por un conjunto de fotodiodos¹, que sustituyen las teselas y recogen la imagen que sobre él proyecta el objetivo, y que después, es guardada como imagen digital. Esta imagen está compuesta por un número de puntos/píxeles² de luz igual al número de fotodiodos de que se compone el sensor.

1.2 Resolución.

En la técnica digital resolución es la capacidad que tiene un dispositivo para captar y reproducir los detalles más pequeños de una imagen. La resolución de una imagen está directamente relacionada con el número de píxeles que proporciona el dispositivo que la crea.

¹ Fotodiodo: es un componente electrónico sensible a la luz que genera una señal de salida de corriente, de una magnitud correspondiente a la cantidad de luz recibida.

² Píxel : Es el elemento más pequeño de que se compone una imagen. La señal que genera un fotodiodo del sensor, crea un píxel en la imagen final que se visualiza en la pantalla del ordenador o en el papel impreso. La palabra píxel se suele aplicar indistintamente tanto al fotodiodo del sensor, como al punto que forma la imagen

A mayor número de píxeles, más resolución de imagen.

Se acepta como norma (pendiente de implementación) de Resolución para la digitalización, que la imagen esté formada con la cantidad de píxeles necesarios para ser reproducida en papel con calidad fotográfica³ a su tamaño original.

1.3 Profundidad de color.

Indica el número de tonos o de colores que puede presentar una imagen. Se regula por la cantidad de bits⁴ que se añaden a cada píxel. Las más usuales son :

- 1 bit : Blanco y Negro. Solamente muestra estos dos colores. Para formar la imagen se toma en cuenta un tono de la misma (es posible seleccionarlo) y a partir del mismo, todos los tonos hacia abajo se traducen al negro, y hacia arriba al blanco, desapareciendo los tonos intermedios.
- 8 bits : Todos los tonos o colores se traducen a una escala de 256 tonos de grises o de colores.
- 24 bits : Llamado Color Real. La imagen puede presentar 16,7 millones de colores. Es la máxima capacidad de percepción que tiene el ojo humano.

1.4 Formatos de archivo / Compresión.

Para conservar la imagen digitalizada debe utilizarse un formato de archivo informático. En el mercado existe una gran variedad de ellos, cada uno está diseñado con unas prestaciones específicas, por tanto, deberá elegirse los que ofrezcan una mayor idoneidad al objetivo final, y a la vez, universalidad de uso para facilitar el acceso a todos los usuarios.

Antes de pasar a describir los más utilizados debe conocerse una técnica llamada Compresión, que permite reducir el tamaño informático del archivo para lograr ahorro de espacio en el soporte de almacenamiento y mayor agilidad de manejo en el proceso informático.

- Compresión : Es una técnica que permite abreviar matemáticamente, por medio de algoritmos muy elaborados, la cadena de código binario de la imagen. La compresión puede ser :

Sin pérdidas : abrevian el código binario sin desechar información, cuando se descomprime la imagen se reproduce bit a bit como la original. No se consiguen grandes reducciones de tamaño de archivo.

³ Para que una imagen pueda imprimirse con calidad fotográfica, se exige que su resolución sea de entre 256 y 300 ppp (puntos por pulgada).

⁴ Bit : unidad mínima de información utilizada en informática. Permite añadir datos .

Con pérdidas : este sistema desecha información. Utiliza unos códigos para eliminar la información menos relevante basándose en la percepción del ojo humano. Esta técnica ha sido optimizada de modo que su resultado puede considerarse como una imagen “sin pérdida visual”. Un documento tamaño DIN A4 en un formato sin comprimir puede tener un tamaño informático de 25 MB y, en un formato comprimido puede reducirse a 2 MB, y será difícil distinguirlos visualmente.

- Formatos de archivo : los más usuales son :

TIFF : Conserva la información original completa, acepta la compresión sin pérdidas, pero no disminuye su tamaño significativamente. No es útil como archivo de difusión debido al gran tamaño de archivo que genera, se suele utilizar como archivo de conservación, sobre todo en fotografía.

JPEG : Admite compresión con pérdidas. Con mínimos niveles de compresión es adecuado para archivo de conservación. Su capacidad de permitir distintas tasas de compresión lo hace muy adecuado para archivo de difusión. Es un estándar muy extendido y mantiene una calidad de imagen muy buena.

PDF : De uso muy extendido en Administraciones, Universidades, Web, etc. Como el JPEG permite distintos grados de compresión, también con pérdidas. Recomendable para difusión. Es perfecto para crear documentos secuenciales con páginas múltiples.

RAW : opción en las cámaras digitales. Guarda toda la información recogida por el sensor, pero, no es un formato estándar, cada fabricante tiene un sistema propietario, no compatible entre ellos y, no es manejable por sí mismo, es decir, aún cuando es posible editarlo, para su uso posterior debe ser convertido a otro formato (Tiff, Jpeg, etc.).

En fotografía se considera el formato RAW como el negativo digital perfecto, ya que recoge la información generada por el sensor sin manipulación alguna, el fotógrafo, más tarde decidirá los parámetros que se aplicarán a la imagen. Para tratar las imágenes individualmente es perfecto. No se aconseja como formato de conservación debido a no ser un formato estándar compatible.

1.5. Soportes/media para archivos digitales.

Para la elección del soporte se tendrá en cuenta la garantía de preservación de la información y su facilidad de reproducción. Los posibles soportes son :

1. DVD-Rom (4,4 GB de capacidad, de doble capa, 8,2 GB), muy utilizado en la distribución de datos. Sustituye al CD-Rom por su mayor capacidad. Una vez grabado no es posible alterar la información contenida, lo que le hace muy adecuado como soporte de seguridad contra manipulaciones de contenido. Su duplicación es sencilla y barata, aunque algo lenta. Son objeto de crítica por la dudosa longevidad que se les supone, para ser utilizados como soporte de archivo de conservación. Algunos fabricantes prometen cien años de vida útil, lo que, obviamente, está por confirmar. El acceso a la información es más lenta que en otros soportes.

2. DVD BlueRay (40/50 GB de capacidad), es un soporte que lleva algún tiempo en el mercado, pero por razones ignoradas no acaba de imponerse y ser aceptado por el público. Se compone de los mismos materiales que el DVD-Rom, por lo que adolece tanto de sus virtudes como de sus defectos.

3. DD (Disco duro mecánico exterior, hasta 2TB de capacidad), gran capacidad de almacenamiento, alta velocidad de acceso a la información, bajo costo de almacenaje, duplicación rápida y barata.

Por contra, es de manejo frágil al llevar incorporada una mecánica que puede sufrir roturas por golpes o caídas, anulando su contenido. También puede verse afectado con pérdidas de información si se acerca a fuertes campos magnéticos. Y muy importante, su contenido puede ser manipulado.

4. Llave de memoria (hasta 64 GB de capacidad), está construida con memorias flash estáticas, esto implica que no tiene partes mecánicas ni elementos en movimiento. No le afectan los campos magnéticos ni las caídas o golpes a no ser que destruyan el material físicamente. No necesita alimentación eléctrica. Permite un acceso instantáneo a la información y su duplicación es fácil y rápida.

Es de momento un soporte caro para su capacidad, pero cada día va aumentando la misma y abaratando su precio. Por contra su contenido puede ser manipulado.

5. SSD (Disco duro estático, hasta 2 TB de capacidad), el aspecto exterior es parecido a disco duro mecánico, pero interiormente está formado por memorias flash de tecnología similar a la Llave de memoria, por tanto ofrece sus mismas virtudes y defectos.

El coste de almacenamiento es muy caro comparado con el de los discos duros mecánicos. Los datos también puede ser manipulados.

(De aquí en adelante cuando se hable de un Disco duro se referirá indistintamente a un disco duro mecánico o al tipo SSD de memorias estáticas).

Recomendaciones : Para la creación del Registro/Archivo digital original y su conservación se recomienda el DVD por las siguientes razones .

- se trata de un soporte barato, fiable y estándar
- puede ser leído por cualquier sistema operativo (Mac, Windows, Linux, Unix)
- no permite que la imagen o información sea manipulada o adulterada una vez grabado el DVD.
- la obtención de copias es barata y sencilla.

2. Tamaños y posicionamiento de la documentación en la toma de imagen.

-Se toma como base de partida para el encuadre de la cámara digitalizadora la de un documento encuadernado de tamaño legal⁵ en su posición abierta, presentando dos páginas a la vista.

-Por tanto se ajustará la cámara para captar en un fotograma la totalidad de las dos páginas, dejando un borde alrededor del documento como comprobación de la visión de su totalidad.

-Los documento de más de un folio (cuadernos, cosidos o encuadernados), se digitalizaran a doble página, es obvio que la primera y la última serán de una página. Los demás documentos de tamaños menores se posicionaran dentro del mismo encuadre, sin variar la posición de la cámara⁶.

La razón de este modo de actuar corresponde a :

-Permitir al usuario que perciba de manera inmediata y a simple vista el distinto tamaño entre documentos.

-Evitar que el operario pierda el tiempo ajustando el encuadre de la cámara a cada tamaño e indicando la reducción utilizada en cada caso.

En caso de bloques extensos de documentos de menor tamaño, puede ajustarse el encuadre indicando la reducción aplicada.

⁵ La documentación antigua no tiene un formato estándar preciso. El folio normal suele llamarse “legal” y tiene un tamaño ligeramente mayor que el estandarizado DIN A4, existe el medio folio y documentos sueltos de diversos tamaños.

⁶ Por una apreciación estética en ocasiones se suelen recortar los márgenes no utilizados en el fotograma alrededor del documento.

Existen tres soluciones:

1 – que cada documento sea encuadrado en la cámara dejando el mínimo margen posible. Esto ahorrará el trabajo posterior de recortar los bordes sobrantes, pero ralentizará la digitalización al obligar a ajustar continuamente la cámara, e indicar en cada fotograma la reducción aplicada.

2 – Programar el ordenador con unas acciones automáticas que permiten manipular ficheros enteros, para quitar márgenes. No es recomendable, porque las acciones son automáticas pero no infalibles. Según el tono y la regularidad del perímetro del documento, puede resultar cortada y anulada parte del mismo, esto exigiría revisar documento a documento todo el fichero y corregir todos los fallos ocasionados.

3 – Recorte manual. Se abre el fichero en un programa de edición de imágenes, se recortan los márgenes y se vuelve a guardar recortado. Es muy lento y se pierde el factor de reducción aplicado al documento, y cuando el usuario vaya a consultar la documentación perderá totalmente la noción del tamaño de un documento en relación con los demás.

Curso de verano de la Universidad de Navarra.

La Hoja de ruta de los archivos personales, familiares y de empresa

Se recomienda utilizar dos reducciones :

a) para documentos de DIN A5 doble y menores.

b) para documentos mayores de DIN A5 doble, hasta tamaño “legal”.

-Cada documento es único, por tanto se digitalizará en solitario. No se agruparan distintos documentos para aprovechar el fotograma.

-La posición de la documentación a la hora de digitalizarse debe ser la posición de lectura del documento.

-Con el fin de facilitar la lectura, los documentos se posicionarán centrándolos en el fotograma en la parte superior del mismo.

-En documentos sueltos se darán casos en el que el documento sea solo de una hoja y escrito por una cara. Se recomienda digitalizarlo así, sin encadenar la imagen con el siguiente.

-En los documentos encuadernados es frecuente encontrar páginas en posiciones invertidas, en estos casos, se digitalizará tal cual aparece en el documento, a la hora de consultarlo no hay problema en el ordenador para girar la imagen y el usuario percibe la posición real del documento.

3. Creación del Registro/Archivo digital

La digitalización tiene como primer objetivo la creación de un Registro o Archivo en soporte digital, que está compuesto por la secuencia de páginas de que consta el documento o expediente original.

Tiene dos funciones :

-conservación permanente de los archivos originales digitalizados

-generar copias de difusión

El archivo de conservación y seguridad será creado con el nivel de calidad suficiente para cubrir los procesos posteriores a los que se vaya a someter. A partir de este archivo se puede generar archivos posteriores de menor calidad, teniéndose en cuenta que será imposible mejorarla.

El debate en torno a esta cuestión puede discutirse en tres puntos :

Calidad	Pros	Contras
Digitalizar a la mayor calidad posible.	A priori parece la mejor solución. Se genera un archivo de seguridad de la mayor calidad posible que permitirá crear archivos menores para otras opciones de uso.	Las posibilidades de equipos digitalizadores de antaño no alcanzaban la calidad que hoy en día nos parecen mínimas. La calidad de hoy estará obsoleta mañana. ¿Dónde está el límite? ¿Se actualizará día a día los equipos para lograr el máximo posible de calidad?
Subordinar la calidad a los usos posteriores	Si el objetivo es únicamente la lectura en pantalla o en la WEB, ¿Porqué generar archivos de mayor calidad? Los archivos digitales serán muy pequeños y su manejo informático será muy ágil. Además su conservación ocupará muy poco espacio.	La informática y sus equipos se supera día a día. Es posible que en el futuro se puedan manejar archivos de mayor calidad y tamaño sin perder prestaciones. Si no se poseen archivos de mayor calidad es imposible crearlos a partir de archivos menores. Ello obligaría a digitalizarlos de nuevo.
El documento se digitaliza con una calidad y resolución que permita su reproducción impresa en papel al tamaño del original y con calidad fotográfica.	Se crea un archivo adecuado para conservarlo como archivo de seguridad y poder reproducir el documento impreso exacto al original. Permite generar otros archivos de menor tamaño para otras opciones como difusión en distintos soportes digitales y en la WEB.	Si no es necesario para algún proyecto en concreto, el esfuerzo y gasto necesario en el equipo y el costo del almacenaje y conservación de tanta información puede parecer innecesario. Sin embargo, la función primordial de crear un archivo de conservación que evite posteriores digitalizaciones, justifica la creación de un Registro de archivo de estas características.

Se recomienda la tercera posibilidad porque :

-Un nivel de calidad alto permite generar archivos de calidades inferiores, mientras que el proceso inverso es imposible.

-Junto con el criterio de conservación debe contemplarse la difusión, por lo que uniendo ambos criterios (Conservación + difusión) se asumirá un matiz muy importante : se guardará con la calidad suficiente.

-Se recomienda utilizar un formato de archivo estándar y de libre uso. El formato JPEG es el más recomendable. Se trata de un formato muy optimizado y utilizado universalmente.

3.1. Recomendaciones para la creación del Registro/Archivo digital original como archivo de seguridad y conservación.

Con el fin de conservar y reproducir con la mayor fidelidad posible tanto su contenido como su estado físico, los documentos manuscritos de archivo de valor histórico se digitalizarán a color real con una profundidad de color de 24 bits.

La resolución digital mínima será la necesaria para poder imprimir en papel el documento, con calidad fotográfica a tamaño del documento original. (Entre 256 y 300 ppp.).

Se guardará el archivo original de salida de cámara directamente, en formato JPEG y con la mínima tasa de compresión.

El archivo Digital Original se considera un Registro de valor permanente y se guardará una copia en DVD y otra en Disco duro.

El formato de grabación del DVD y disco duro será híbrido, es decir, capaz de ser leído por cualquier plataforma informática (Mac, Windows, Linux, Unix). Una vez grabados el DVD y Disco duro se controlará que los archivos se abren correctamente y se testeará su contenido.

Ambos, DVD y DD serán guardados en diferentes ubicaciones y el almacenamiento deberá reunir las características ambientales recomendadas por el fabricante.

Se recomienda no manipular ni tocar la superficie de lectura del DVD, no se colocará ni se imprimirá etiqueta alguna sobre él. Se guardará en la caja suministrada por el fabricante, eliminando todas las carátulas de origen y solamente se añadirá, en la caja, una etiqueta con el código de identificación del DVD.

El disco duro será del tipo de los alimentados directamente del ordenador para eliminar elementos susceptibles de sufrir averías. Se etiquetará con su código correspondiente.

Del disco duro antes mencionado se creará una copia, también en disco duro, que se denominará Master, y que servirá para generar todas las copias futuras que sean necesarias.

3.2. Recomendaciones para la creación de copias de difusión.

En la copia de difusión no es necesario conservar los niveles de calidad del archivo de conservación. Es suficiente con que permita una lectura perfecta en pantalla y

una impresión en papel en el que se pueda leer su contenido con claridad. Existen varias posibilidades :

-generar archivos en formato JPEG con un tamaño entre 1,5/ 2 MB por fotograma.

generar archivos en formato PDF, con un tamaño entre 1,5/2 MB por fotograma. Este formato permite crear PDF multipágina (todas las páginas de un documento se juntan en un archivo único a modo de libro que reproducen fielmente la secuencia de páginas del original), facilitando así la consulta de la documentación.

-descartando el DVD como soporte de difusión, debido a que fragmenta el fondo, y sustituyéndolo por la Llave de memoria o el disco duro que permiten contener la totalidad del mismo, puede pensarse en incluir una base de datos que permita realizar búsquedas y enlaces directos a todos los documentos del mismo.

Como se ha apuntado antes no se recomienda el DVD como soporte de difusión porque :

-Aunque el soporte sea más barato, el tiempo empleado en la grabación y etiquetado del DVD y de la caja que lo contiene, encarece el producto a niveles de los otros soportes.

-No permite un acceso directo instantáneo a todo el fondo, debido a su capacidad debe fragmentarse y esto dificulta y retarda la consulta del archivo

-El acceso a la lectura en el DVD es lento comparado con los otros soportes y se ralentiza más debido al trasiego de DVDs para acceder a todo el fondo.

Otro aspecto a tener en cuenta en relación a las copias de difusión es el control de uso sobre las mismas, que desee mantener, ya sea el propietario o el Organismo que edite las copias de archivo.

Si se desea controlar el uso del fondo, debe descartarse difundir el mismo, por medio cualquier soporte/media autónoma.

Un fondo de archivo histórico, entregado a un usuario en una llave de memoria, en un disco duro, o en una colección de DVD, puede en un plazo mínimo de tiempo aparecer duplicado miles de veces, e incluso colgarse en una página Web con descarga libre.

Contra esta situación, los Organismos, Universidades o entidades que vayan a permitir consultar o difundir sus fondos, pero quieran controlar su uso, solo cabe pensar en una consulta del archivo por red (ya sea Internet o corporativa), controlando los accesos y no permitiendo bajadas de documentos sin control.

Resumen a vista rápida :

Copias	Formato	Soporte	Destino / Uso
Registro digital Original	JPEG Archivo generado directamente desde cámara digitalizadora. Tamaño 10/12 MB por fotograma	DVD, como soporte de conservación y seguridad, y Disco Duro como soporte para generar un Master del que se crearan las copias futuras	DVD : archivo original de conservación indefinida. Controles periódicos de conservación. Objetivo : Conservar una copia sobre un soporte que no permite la alteración de su contenido, para preservar la integridad y autenticidad de la digitalización. Aunque se le supone una vida de entre 30 y 100 años, esta duración no está garantizada por lo que es necesario pensar en su replicación sobre un soporte de iguales o mejores características. No debe utilizarse a excepción de ser la única copia existente para generar otra, o migrar a otro soporte. Disco duro : archivo para generar la copia Master del siguiente apartado. No debe ser utilizado más que para esta función. Ambos deberán ser guardados en distintas ubicaciones, en sitio seguro y en las condiciones ambientales recomendadas por el fabricante.
Master	JPEG	Disco duro	Copia del Disco duro del archivo original. De él se obtienen todas las copias posteriores. No se utiliza más que para esta misión.
Copias de difusión/uso	JPEG, PDF en modo de multipágina y otros	DVD Llave memoria Disco duro	Distribuidas a propietarios y usuarios. Si es necesario pueden utilizarse otros formatos de archivo. No aconsejable si se quiere controlar el uso del fondo. (Ver texto, Creación de copias de difusión, 3.2).
Difusión en Internet	JPEG	Servidor	Difusión on-line Se busca la arquitectura informática más adecuada para el acceso al usuario. La más adecuada para controlar los accesos, consultas y descargas, como control de uso y utilización del fondo.

4. Fases del proceso de digitalización y control de calidad.

-Recepción del cuaderno-guía. El cuaderno-guía es la base de datos impresa del fondo de archivo a digitalizar. Además de la descripción de cada documento o ítem, debe incluir el código de identificación del archivo digital y el código de cada documento⁷, además, un campo de observaciones en el que el operario anotará las incidencias. Con éstas se crea un informe de digitalización final.

-Recepción y manipulación de la documentación. La documentación debe estar ordenada de acuerdo con el cuaderno-guía. Como quiera que cada documento puede estar compuesto por un número indefinido de páginas y éstas pueden no estar numeradas, se respetará el orden en que se reciban.

-Digitalización. Creación del Archivo Digital original, según las normas descritas en 3 y 3.1.

-Creación de archivos de conservación y difusión, según las normas descritas en 3.2.

-Control de calidad. Al final del proceso deberá efectuarse un control de calidad, que será la verificación de que los archivos grabados se abren y leen correctamente en los soportes en que se hallan guardados.

-Se establecerá un programa de control de conservación de los soportes para tomar las medidas oportunas ante la eventualidad de su posible deterioro.

Además de este control físico, deberá preverse la inutilización del contenido por cambios informáticos. Los avances en hardware+software pueden dejar obsoletos los soportes anulando la posibilidad de su consulta, por lo que debe diseñarse un plan de migración.

Donostia Setiembre 2010

⁷ Ejemplo de código : **0770001100078**, siendo :

007 : código de colección o fondo de archivo,

00011 : código del documento o legajo, y

00078 : código de la página del documento o legajo.